Imperfect Rhymes as a Measure of Phonological Similarity
Aaron Kaplan & John Woodmansee
University of Utah & NCSSM
a.kaplan@utah.edu
UNC Spring Colloquium
April 7, 2018

1 Similarity in Phonology
• The degree of similarity between segments is central to many domains of phonology:
  – IO & BR Faithfulness (McCarthy & Prince 1995)
  – OCP (e.g. Leben 1973): adjacent elements must be dissimilar
  – Agreement by Correspondence (e.g. Rose & Walker 2004): harmony between segments that meet a threshold of similarity
• Intuition: speakers are aware of and can measure how similar segments are. Sometimes similarity is avoided (OCP), and sometimes it is reinforced (ABC)
• Similarity is measured with distinctive features, and all features are equal.
• Does this match speakers’ intuitions about similarity?
⇒ Do more featural differences = greater dissimilarity?
⇒ Is a difference in [±F] equivalent to a difference in [±G]?

2 Imperfect Rhymes
• Imperfect rhymes: sometimes rhyming words don’t rhyme exactly:
(1) This version of the world will not be here long [lag]
It is already gone It is already gone [gan] T Bone Burnett, “Palestine, Texas”
• Assuming lyricists are more likely to use similar-sounding imperfect rhymes than dissimilar ones, we can use imperfect rhymes to probe speakers’ judgments about segmental similarity.
• If featural similarity matches speakers’ judgments about similarity, the frequency of consonantal pairings in imperfect rhymes should be inversely proportional to the number of features they mismatch on.

3 Our Study
3.1 The Data
• Our study: rhymes from 117 songs from many genres of popular music; 1977–2016.
• Data collected by AK and students at the North Carolina School of Science and Mathematics.
  – Juniors in John Woodmansee & Ormand Moore’s 2016–2017 American Studies class
• For today, 294 rhyming pairs of words meeting the following criteria:
  – “Masculine” rhymes: the stressed/rhyming syllables are final: unfair/compare
  – “Feminine” rhymes (treble/rebel): stressed syllable and all following syllables “should” match. Not sure how to handle them yet...
  – Identical vowels (analysis here focuses on consonants)
  – Same number of consonants: long/gone but not fun/fund
• Identical pairs included unless the pair is repeated in identical lines (e.g. it’s in the chorus).
• Transcriptions pulled from CMU Pronouncing Dictionary
• In two words with shape . . . VC_{C_1}C_{C_2} . . . C_{C_n}, we compared C_{C_1} to C_{C_1}, C_{C_2} to C_{C_2}, etc.
  – This doesn’t account for cases where Word 1’s C_{C_1} matches Word 2’s C_{C_2}, but it’s a good first approximation.
• Total: 378 pairs of mismatched consonants
3.2 Evaluating Featural Similarity

- Our feature system: an “average” of commonly accepted systems, perhaps most similar to Hayes (2009).
- Uncontroversial features: [syll, son, approx, voc (= cons), lat, nas, cont, voi]
- [delayed release] to distinguish stops from affricates (fricatives are [−d.r.], contra Hayes)
- Place features: to avoid inflation of featural differences, we used [lab, dental, cor, pal, dor] instead of [lab, cor, dor] with many dependent place features.
- This idealized feature system provides a rough starting point: do distinctive feature systems in general have a hope of reflecting speakers’ judgments?

4 Results

4.1 General Trends

- Most common consonant pairs:

(2)

<table>
<thead>
<tr>
<th>Consonant Pairs</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>n &gt; 5</td>
<td></td>
</tr>
</tbody>
</table>

- Pairs with fewer featural differences more common, for the most part:

(3)

<table>
<thead>
<tr>
<th>Number of Features on which Consonants Differ</th>
<th>Number of Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>66</td>
</tr>
<tr>
<td>2</td>
<td>149</td>
</tr>
<tr>
<td>3</td>
<td>73</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

- The numbers:
  - One feature different: 66
  - Two features different: 149
  - Three features different: 73
  - Four features different: 55
  - Five features different: 25
  - Six features different: 9
    - smile/tune × 2; while/tune (Colbie Caillat, “Bubbly”)
    - whole/home; close/home × 2; nine/life × 2 (Eminem, “Lose Yourself”)
    - roof/moon (Tom Petty, “Even the Losers”)
  - Seven features different: 1
    - whole/broke (Eminem, “Lose Yourself”)
• Low number of 1-feature differences: caused by place features
• A multivalued [Place] feature smooths things out:

Number of Consonant pair Mismatches by Number of Featural Differences using Multivalued [place]

![Graph showing mismatched features with just 1 feature difference.](image)

• And with just 3 features:

Number of Consonant pair Mismatches by Number of Featural Differences using Place, Manner, and Voice

![Graph showing mismatched features with just 1 feature difference.](image)

• These simplifications suggest that features and speakers’ judgments are related.

4.2 Not All Features are Equal
• If exactly one feature mismatches:

Mismatched Features with just 1 Feature Difference

![Graph showing mismatched features with just 1 feature difference.](image)

• Interim Summary
  – Distinctive features do a decent job of modeling imperfect rhyme frequency.
  – Featural differences match speakers’ similarity intuitions . . .
  – Except for place features: mismatches in place mean a large number of featural differences, but this is not reflected in the frequency of pairs mismatching in place.
  – Fewer multivalued features perform better than many binary features.
  – For the future: compare specific feature systems.
• If exactly two features mismatch:

(8)

![Graph showing mismatched features with just 2 feature differences](image)

- Mismatched Features with just 2 Feature Differences
- Frequency
- Feature Pairs
- Mismatched Features with just 2 Feature Differences

• Some mismatch more than others.

• To ensure this isn’t simply a reflection of consonantal frequency, we did the same analysis on the portion of the CMU dictionary that also occurs in CELEX (Baayen et al. 1995) to weed out low-frequency items:
  - Match each final-stress word to all other words with the same final vowel and same number of consonants
  - Compare coda consonants as before

(9)

![Feature Mismatches in Song Rhymes and CMU](image)

* = \( p < .05 \) (Bonferroni correction)

- Over represented: [lab, cor]
- Under represented: nearly everything else
- Mismatches on [lab, cor] are more acceptable. Perhaps differences along these dimensions are “smaller” than differences along other dimensions.

- What’s up with [lab] & [cor]?
  - [m] \( \sim \) [n]: 31.1% (60/193) of all [lab] mismatches; 27.8% (60/216) of [cor] mismatches.
  - This accounts entirely for the prevalence of [lab] and [cor] mismatches.
  - We can’t explain the high frequency of [m] \( \sim \) [n] merely on the grounds that place cues for nasals are weak: why are [m] \( \sim \) [n] and [n] \( \sim \) [h] infrequent?
    - 9 tokens of [n] \( \sim \) [h]: 18.4% of [dor] mismatches, 4.2% of [cor] mismatches
    - 1 token of [m] \( \sim \) [h]: 2.0% of [dor] mismatches, .5% of [lab] mismatches
  - It looks like a combination of nasal place weakness and a preference for [lab]/[cor].

- What this might mean:
  - Certain feature (mis)matches are more significant than others, as are certain combinations.
  - E.g. labials and coronals are judged as more similar than, say, labials and dentals, stops and fricatives, etc.
  - If featural asymmetries matter to grammars, they should arise in the typology of ABC/OCP systems.
    - Cooccurrence of similar consonants is disfavored in \( C_1C_2C_3 \) Arabic roots. Frisch et al. (2004): all combinations of non-identical place features in \( C_1 \) and \( C_3 \) are over represented, but labial/dorsal combinations are less over represented than others.
    - Not so for \( C_1 \) and \( C_2 \) though
  - But maybe grammars don’t care about these asymmetries. Grammars are a step removed from phonetic detail in other ways.
Comparison with Zwicky (1976)

- Most Common Consonant Pairs

\[(10) \text{a.} \]

<table>
<thead>
<tr>
<th>Pair</th>
<th>Count</th>
<th>In Zwicky?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

\[(11) \]

- Zwicky’s (1976) results (for feature mismatches ≥ 10):
  - [dor] 148
  - [lab] 138
  - [cor] 70
  - [cont] 49
  - [voi] 19
  - [pal] 10

Conclusion

- Generally, fewer featural differences between consonants makes them more likely to be paired in rhymes.
- Except for place features, counting features is a plausible model of speakers’ similarity judgments.
- But the particular features involved matters, too: do some they represent smaller differences?
- Next Steps
  - Vowels
  - Differences in number of consonants
– Compare specific feature systems
– Morphology (Zwicky 1976): e.g. does past-tense /d/ behave differently from other /d/?
– Genre & year differences

References


